ELSEVIER

Journal of Food Composition and Analysis 19 (2006) S86–S90

JOURNAL OF FOOD COMPOSITION AND ANALYSIS

www.elsevier.com/locate/jfca

Report

Quality assurance during data processing of food and nutrient intakes

Jaswinder Anand*, Nancy R. Raper, Amy Tong

Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, USDA, Beltsville, MD 20705, USA

Received 11 February 2005; received in revised form 14 February 2006; accepted 24 February 2006

Abstract

The Food Surveys Research Group (FSRG) is responsible for methods of data collection and processing of dietary intake data, including the *What We Eat in America* Survey, which is the dietary interview component of the National Health and Nutrition Examination Survey (NHANES). Many measures have been implemented to ensure accuracy of the dietary data, such as a state-of-the-art data collection instrument and an extensive food and nutrient database. Also important, but usually receiving less public attention is the quality assurance taking place during data processing. A four-part quality assurance procedure is used for food intake data processed by FSRG. This includes (1) determination of overall acceptability of each 24-h recall, (2) resolution of new or unusual foods and amounts, (3) administration of data integrity checks, and (4) verification of 24-h recalls with extremely high nutrient intakes. Quality assurance not only contributes to the accuracy and validity of a specific set of dietary intake data, it also benefits future studies because findings help identify areas to target for interviewer training or for improvements in data collection and processing procedures.
Published by Elsevier Inc.

Keywords: Food and nutrient intakes; Quality control

1. Introduction

The Food Surveys Research Group (FSRG) is part of the Beltsville Human Nutrition Research Center at the US Department of Agriculture (USDA). Its mission is to monitor and assess food consumption and related behavior of the US population by conducting surveys and providing the resulting information for food and nutrition-related programs and public policy decisions. USDA has studied what Americans eat for over a century and began conducting nationwide surveys in 1935 (Tippett et al., 1999). The data from these surveys are used by government, industry and academia for many purposes (Tippett and Cypel, 1997). In recent years, the leadership of USDA and the Department of Health and Human Services (HHS) has integrated their efforts to collect national food consumption data (McDowell, 2003). Since January 2002, FSRG has provided the food intake methodology and processing system for the dietary interview component of the HHS National Health and Nutrition Examination Survey (NHANES). This dietary component is called *What We Eat in America*. Approximately, 10,000 24-h dietary recalls are processed at FSRG each year as part of this responsibility. Additional dietary data are also processed for other studies conducted by FSRG. The USDA Food and Nutrient Database for Dietary Studies (FNDDS) is used to determine nutrient intakes (USDA, 2004). The source of nutrient data is the USDA National Nutrient Data Base for Standard Reference (USDA, 2005), which includes documentation about the derivation (analytical or calculated) for each nutrient value.

A comprehensive and effective quality assurance program is an important component of any survey data processing system. The quality assurance activities at FSRG are built on several decades of experience in identifying and eliminating the source of errors and inconsistencies that may occur in food and nutrient intake data. Fortunately, automated collection and coding methods have eliminated some past sources of errors, although quality assurance remains a priority for those operations also. Briefly, the data collection instrument, USDA's Automated Multiple Pass Method (AMPM), provides for standardized and complete data collection,

*Corresponding author. Tel.: +1 301 504 0181; fax: +1 301 504 0377.
E-mail address: janand@rbhnrc.usda.gov (J. Anand).

0889-1575/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jfca.2006.02.005

Table 1
Automated Multiple Pass Method (AMPM)

Step	Pass	Purpose
1	Quick list	To collect a list of foods consumed the previous day
2	Forgotten foods list	To collect foods that may have been forgotten during the quick list; questions probe for foods by categories: nonalcoholic beverages; alcoholic beverages; sweets; savory snacks; fruits, vegetables, cheese; breads and rolls; other foods
3	Time and occasion	To collect time and name of eating occasion for each food; used to sort foods chronologically and group into eating occasions
4	Detail and review	To collect a detailed description of each food consumed, including amount eaten and additions to the food; also, to review eating occasions and times between occasions to elicit forgotten foods
5	Final review	To collect additional foods not remembered earlier

using a 5-step interview outlined in Table 1. The software ensures that appropriate questions are asked, and built-in edits guard against implausible responses (Raper et al., 2004). Codes are assigned to variables other than foods and amounts, e.g. source of food within the AMPM. For foods and amounts, codes are assigned in a 2-step procedure using automatic coding techniques (codes assigned by computer) for commonly reported foods and computer-assisted coding (codes assigned by individuals) for the remainder. For the computer-assisted food coding part, Survey Net (Raper et al., 2004), a system developed and perfected for the Continuing Survey of Food Intakes by Individuals (CSFII) (USDA, 2000), continues to be used for *What We Eat in America* and other FSRG studies.

What We Eat in America data are received at FSRG after data collection and food coding have already taken place. Comprehensive training and monitoring programs for both interviewers and coders complement the automated systems to help ensure that data are of good quality. Nevertheless, data continue to undergo the standard FSRG quality assurance procedures as data processing is completed. The remainder of this paper describes those procedures.

2. Methods

Coded 24-h dietary recalls received at FSRG for data processing undergo four quality assurance steps as the food intake data are edited and then analyzed for their nutritional content. These steps include an evaluation of the overall acceptability of each 24-h recall, resolution or confirmation of new or unusual foods and amounts reported by respondents, administration of data integrity checks, and verification of 24-h recalls with extremely high nutrient intakes. Each step is discussed below.

2.1. Determination of overall acceptability of each 24-h recall

Survey net, mentioned above as being used for computer-assisted food coding, is also used at FSRG for on-line review and editing of the 24-h recalls, as well as for nutrient analysis of the recalls. One of the first quality assurance activities is to confirm that each 24-h food recall meets established minimum criteria for completeness and that it contains reliable data. Food recalls that either fail to meet minimum criteria or do not qualify as reliable will be excluded from future data analyses, such as determination of mean nutrient intakes for the population. The minimum criteria required for intakes collected with the AMPM are (1) that the first four steps of the 5-step AMPM were completed during the interview, and (2) that the intake contains no "missing foods". Failure to meet the first condition happens infrequently; but it results if the respondent stops the interview before step 4 is completed. Step 4 is critical because specific details about foods and amounts are collected at this point, and a number of questions that probe for additional foods are also administered at this time. The second condition, "missing foods", occurs when an adult serving as a proxy, usually for a child, has reported that the sample person ate a meal, usually at day-care or school, but does not know what was served or how much was consumed. Later, after obtaining permission from the parent or guardian, an interviewer contacts the day-care provider, school, or other source of the meal, to collect the missing information. When this attempt, called "data retrieval", is not successful, the recall is marked as failing to meet minimum criteria. Coded 24-h recalls for *What We Eat in America* are received electronically once a month at FSRG with the preliminary determinations regarding minimum criteria already made. FSRG nutritionists review recalls marked as failing minimum criteria to confirm their status.

It is possible for a recall to meet minimum criteria, but still not be acceptable for use in further research. Interviewers assess whether or not they believe each 24-h recall is reliable based on their knowledge of how the interview transpired, and they provide reasons whenever an assessment of "unreliable" is made. Reasons vary, but "unreliable" may result if a sample person has been uncooperative, or could not fully participate because of memory or other difficulties and a suitable proxy was not available. Each 24-h recall marked as unreliable is reviewed by survey nutritionists at FSRG who judge the usability of the data.

2.2. Resolution or confirmation of new or unusual foods and amounts reported by respondents

Foods and amounts reported in *What We Eat in America* and other FSRG dietary studies are coded using the FNDDS. This database, formerly called the Survey Nutrient Database, includes descriptions for over 13,000 foods and about 30,000 portion sizes for those foods. Still, some foods and amounts reported by respondents do not match database descriptions because many new products are introduced to the United States market each year. In fact, according to Mintel's New Product Database, 11,500 new food products were introduced in 2003 (IFT Newsletter, 2004). It is not feasible for FSRG to compile information on every new product, but all new foods reported in *What We Eat in America* or other FSRG studies are researched to learn what types of foods they are and whether they fit within existing food code descriptions. For example, a new brand of orange juice drink may fit adequately under one of the existing drink descriptions, such as citrus *juice drink, calcium fortified*, or one of the other citrus drinks. New package sizes reported for existing foods are also researched to confirm that they actually exist and were not mistakenly reported by the respondent (e.g. a 16.9 ounce bottle of juice mistakenly reported as 169 ounces). When a reported food does not match an existing food code or a new size description is not present in the database, Survey Net has special features for handling these situations, marking these items with temporary placeholders called "unknown foods" or "unknown amounts". Once a month, all new cases of unknown foods and unknown amounts are listed from Survey Net and the identification process begins. Information for commercial products is collected over the Internet, from grocery stores, restaurants, or by direct contact with food manufacturers. Because many new products are discontinued after a brief time in the market, new items are not added to the food coding system immediately. New products remain classified as "unknown foods" until they have been reported in multiple 24-h recalls by different respondents, at which time they are added to the database. Otherwise, after a brief holding period, an unknown food is matched and coded to the closest food in the database. Unknown foods are not always new commercial products. Frequently they are mixtures for which a single unique food code is not present in the database. These occurrences are generally resolved by using multiple food codes, linked together with a special code signifying a "combination".

2.3. Administration of data integrity checks

The administration of data integrity checks consists of running the completely coded data through programs written to detect possible errors, which may occur because of misunderstandings by respondents, data entry errors by interviewers, or misinterpretations or keying errors by coders. These checks, which have been based on errors or inconsistencies found in past surveys, have been grouped into three main edit categories—deterministic, query, or fatal edits. These categories, discussed below, are based on classifications developed by Biemer and Lyberg (2003), who have studied and reported on various aspects of survey quality, including techniques for finding and reducing errors in survey data. Guidelines have been developed for each integrity check to facilitate the review when suspicious values are found.

2.3.1. Deterministic edits

Deterministic edits reveal errors that most likely need to be corrected. These edits are conducted to identify foods that may have been coded in the wrong form, such as a concentrated juice that is usually consumed in the reconstituted form. Other foods covered by these checks are dry instant coffee or tea, dry milk, other instant beverage mixes, mashed potato flakes, dry baby formulas, and condensed or dry soups. These foods are generally coded in their reconstituted forms. However, sometimes a code for a dry or concentrated form may be used, accompanied by another code for a liquid food. The two items are linked with a "combination code", to let data users know the foods were mixed together; for example, concentrated lemonade mixed with ice tea. Edit checks find occurrences when dry or concentrated forms of the above-mentioned foods have been coded without a combination code, and they are reviewed according to guidelines for each edit. As an example, the guideline for dry coffee is illustrated in Table 2.

2.3.2. Query edits

The majority of edit checks are classified as query edits, which identify suspicious values that must be reviewed to determine if errors exist. Query edits have been grouped as follows:

Atypical foods for various groups: Several queries check the 24-h recalls for foods which may be unusual for respondents of certain ages. These checks identify alcohol consumed by respondents under the age of eighteen, baby foods consumed by respondents over the age of two, children older than one being nursed, and very young infants consuming foods other than baby food or formula. Each item is verified against the original data collected during the interview and corrections are made if errors are found.

Table 2
Example of an editing guideline

Script name: dilutions_coffee
Identifies: dry coffee code not in combination with a liquid such as water
Procedure:
1. Should the dry coffee have a combination code?
Yes = add combination code
No = continue to next step
2. Is the amount more representative for a liquid form?
Yes = change to liquid coffee
No = code water and link with combination code

Eating occasion edits: Multiple integrity checks are conducted to make sure that meal names and times were coded correctly because these variables are frequently important in secondary research using food survey data. Edit checks identify meals reported at unusual times such as breakfast in the evening or dinner in the morning, and 24-h recalls with these situations are reviewed for possible data collection errors. An edit guideline, similar to the earlier example, aids in making decisions to change values. No changes are made unless mistakes are obvious or coded responses are not credible. For example, a school lunch designated at 12:00 a.m., or a school breakfast at 7:30 p.m., would be changed to 12:00 p.m. and 7:30 a.m., respectively. Reports of breakfast, lunch, or dinner at more than one time of day are also reviewed. Since respondents are encouraged to recall foods in ways that help them remember everything they ate, sometimes they report each food within a meal at different, but close, times. These situations are edited so that all items within the meal reflect the same time to facilitate later use of the data. Other eating occasion edits identify questionable meal names; for example, if an eating occasion is coded as a "drink" and no beverage is reported at this time, the meal name is changed to "snack". Foods consumed at the same time of the day but coded with different meal names are also identified and, all foods are changed to the occasion reported for the majority of the foods.

High and low amounts: Checks for amounts of food consumed are conducted in multiple ways. Amounts are reviewed if they are above or below defined limits. For example, the high limit for milk is 1098 g (4.5 cups) and the low limit is 61 g (0.25 cup). However, small amounts of milk added to coffee or tea would not be reviewed. A frequently reported amount that exceeds the limits for several foods and results in changes is "8 ounces", which is often reported when a respondent actually means the volume measurement of "1 cup". For example, a survey participant might report consuming 8 ounces of a ready-to-eat cereal. In this example, if the respondent were a teenage boy and the amount (227 g or almost half of a cereal box) had been verified by the interviewer, it would be left as originally coded. But if the respondent were a young child and there was no note from the interviewer indicating that the amount had been verified, the quantity would be changed to 1 cup (which is about 1 ounce or 28 g of cereal).

Source of food: Two checks are conducted regarding codes for food source. One check determines if foods coded as "grown or caught by the respondent" are realistic, such as vegetables grown in a garden. The second check identifies restaurant foods for which a respondent has provided specific information about types of fat, milk, or salad dressings that were used in preparation. These situations are reviewed to decide if the specific information is realistic, or if codes representing foods described with less specificity should be used.

2.3.3. Fatal edits

Fatal edits identify inconsistencies which must be corrected for the data to be usable. Although, as the name implies, these errors are the most dangerous, they do not occur often and are actually the easiest to edit. These edit checks locate missing or invalid values for each variable. They also check to ensure that information about each respondent is consistent across multiple days of data. For example, two 24-h recalls for one respondent must contain identical data for variables such as gender or race. All inconsistencies falling into the "fatal edits" category are corrected.

2.4. Verification of 24-h recalls with extremely high nutrient intakes, or outliers

Once the first three steps have been completed for a study, or for one year of *What We Eat in America*, Survey Net calculates preliminary nutrient intakes for the 24-h recalls, and the final quality assurance phase is conducted using the aggregate data. Checks are run separately for the following groups: adult males, adult females, males 12–19, females 12–19, children 6–11 and 1–5, and infants. Within each group, 24-h recalls with daily intakes above the 99th percentile for a nutrient are identified and reviewed to determine the food source of the nutrient. Unusual cases are investigated. For example, when the source of a very high vitamin E intake was identified as *tomato and vegetable juice*, it was found that the database value was based on samples of the food which were fortified with vitamin E. The database value was corrected before the final nutrient analysis. During this quality assurance phase, mean intakes of nutrients and food groups for the above gender/age groups are also compared with intakes from previous years or other similar surveys. Unusually high or low values are investigated.

3. Results and discussion

Although an effective quality assurance program during data processing is time consuming, it is essential for producing accurate and usable food and nutrient intake information. In recent years, this effort has been lightened somewhat by the introduction of automated collection and coding procedures, which have eliminated many of the types of errors once found during data processing. For example, the USDA Automated Multiple Pass Method, used in the integrated *What We Eat in America* since January 2002 and by other studies involving FSRG, minimizes problems such as collection of ambiguous information or insufficient data by guiding the interview through standardized questions about each food reported. Food coding has also been enhanced. Once a potentially error-prone operation, over 50% of foods collected by the AMPM are now coded automatically by a computer program. Still, quality assurance continues to be important throughout data processing. For one thing, it provides the

opportunity to quickly identify deficiencies in either the data collection or coding operations, because decisions about the usability of each 24-h recall takes place shortly after data have been coded. During the next data processing phase, research is conducted to learn about new foods and portion descriptions reported within the recalls. Findings from this work lead to updates in the food and Nutrient Database for Dietary Studies, which is maintained for analysis of 24-h recalls. Automated data integrity checks exist to detect suspicious values in the individual 24-h recalls, which lead to greater accuracy and consistency. Finally, aggregate data are used to identify recalls with the nutrient intake outliers that should be checked.

The value of quality assurance during data processing is two-fold. First, it increases the overall accuracy of a specific set of intake data. Second, it provides an opportunity to identify inconsistencies that might be eliminated by making changes to the automated data collection or coding software. However, it is sometimes simply more efficient to continue finding and correcting types of errors that occur infrequently rather than (a) implementing complicated and expensive modifications to computer software, or (b) increasing respondent burden by lengthening the interview with additional questions that add little value. Also, it is important not to dwell on quality assurance so much that searching for unimportant discrepancies either increases the cost of data processing or delays the release of useful food and nutrient intake information.

Research on editing indicates that not all errors need to be corrected. In fact, in some cases studies have shown that 50% of changes made to data during the editing process resulted in a less than 1% change in final estimates derived from the data (Biemer and Lyberg, 2003). A selective approach to checking and editing data is needed, ensuring that time and resources are spent on correcting errors that affect final estimates derived from the data or that will decrease the usefulness of the data for secondary research.

4. Conclusions

Quality assurance during data processing is an important part of dietary research studies. Its goal should be to provide data of the best quality possible with the time and resources available. FSRG has developed an efficient quality assurance process structured to maximize its effectiveness. It contributes to high-quality food intake data and helps identify changes that may improve data collection, coding, or the food and nutrient database.

References

Biemer, P.P., Lyberg, L.E., 2003. Data Processing: Errors and Their Control. Introduction to Survey Quality. Wiley, New Jersey (402pp.).

Institute of Food Technologist (IFT), 2004. New Product Introductions Exceed 23,000 in 2003. IFT Newsletter, February 25, 2004. Retrieved June 21, 2004 from Institute of Food Technologist Home Page on the World Wide Web: http://www.ift.org/cms/

McDowell, M., 2003. US Department of Health and Human Services—US Department of Agriculture Survey Integration Activities. Journal of Food Composition and Analysis 16, 343–346.

Raper, N., Perloff, B., Ingwersen, L., Steinfeldt, L., Anand, J., 2004. An overview of USDA's Dietary Intake Data System. Journal of Food Composition and Analysis 17, 545–555.

Tippett, K.S., Cypel, Y.S., 1997. Design and Operations: The Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey, 1994–96. Nationwide Food Surveys Report No. 961, US Department of Agriculture, Agricultural Research Service, 197pp. Available at: http://www.ars.usda.gov/sp2userfiles/place/12355000/pdf/Dor9496.pdf. Accessed February 2006.

Tippett, K.S., Enns, C.W., Moshfegh, A.J., 1999. Food Consumption Surveys in the US Department of Agriculture. Nutrition Today 34, 585–597.

United States Department of Agriculture (USDA), 2000. Agricultural Research Service. Continuing Survey of Food Intakes by Individuals 1994–96, 1998. CD-ROM. Accession No. PB2000-500027, National Technical Information Service.

United States Department of Agriculture (USDA), 2004. Agricultural Research Service. USDA Food and Nutrient Database for Dietary Studies, 1.0. Retrieved June 21, 2004 from the Food Surveys Research Group Home Page on the World Wide Web: http://www.ars.usda.gov/ba/bhnrc/fsrg

United States Department of Agriculture (USDA), 2005. Agricultural Research Service. USDA Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Release 18. Retrieved Feb, 2006 from USDA Nutrient Data Laboratory website: http://www.ars.usda.gov/ba/bhnrc/ndl